

KOMPARASI METODE PENYETARAAN VERTIKAL PADA SKOR MATA PELAJARAN MATEMATIKA SISWA SMP NEGERI 1 TIDORE KEPULAUAN

Ramli Salasa, Heri Retnawati
Staff Pengajar STMIK TidoreMandiri, Maluku Utara, Universitas Negeri Yogyakarta
ramlisalasa@ymail.com, retnawati_heriuny@yahoo.co.id

Abstrak

Penelitian ini bertujuan: (1) mengetahui karakteristik butir soal hasil analisis menggunakan teori tes klasik, (2) mengetahui karakteristik butir soal hasil analisis menggunakan teori respon butir model satu parameter (*Model Marginal Maximum Likelihood*), (3) membandingkan hasil analisis butir soal menggunakan teori tes klasik dan teori respon butir, (4) mengetahui kemampuan peserta ujian, dan (5) mengetahui persamaan penyetaraan perangkat tes prestasi belajar Matematika antara kelas VII, VIII, dan IX SMP Negeri 1 Tidore Kepulauan. Penelitian ini dengan metode kuantitatif. Populasi penelitian adalah 606 siswa SMP Negeri 1 Tidore Kepulauan. Estimasi parameter berdasarkan model logistik 1 dan 3 parameter dan menggunakan program BILOG 3.00. Data diolah menggunakan pendekatan teori tes klasik (*Microcat iteman*) dan teori tes modern. Hasil penelitian sudah cukup baik kualitasnya. Berdasarkan tes klasik perangkat tes kelas VII, VIII dan IX sejumlah 29, 27 dan 30 butir, tes modern perangkat tes VII, VIII dan IX sejumlah 28, 27, dan 24 butir. Penyetaraan perangkat tes butir soal dari kelas VII ke kelas VIII adalah $Q^* = 1,084X - 0,221$, dari kelas VIII ke kelas IX adalah $R^* = 0,922X + 2,965$ dan dari kelas VII ke IX adalah $R'^* = 0,998 X + 2,775$. Rata-rata kemampuan peserta ujian kelas masih rendah berkisar 0,265 sampai 0,58. Berdasarkan analisis tes menggunakan dua pendekatan yaitu tes klasik dan tes modern menunjukkan pada ketiga kelas yang diamati terjadi perbedaan persentase validitas butir yang baik.

Kata kunci: *penyetaraan vertikal, tes prestasi belajar, matematika, bank soal*

A COMPARISON OF VERTICAL EQUATING METHODS FOR MATHEMATICS SCORES OF THE STUDENTS OF SMP NEGERI 1 TIDORE KEPULAUAN

Ramli Salasa, Heri Retnawati
Staff Pengajar STMIK TidoreMandiri, Maluku Utara, Universitas Negeri Yogyakarta
ramlisalasa@ymail.com, retnawati_heriuny@yahoo.co.id

Abstract

This study aims to: (1) investigate characteristics of test items based on the results of the analysis using classical test theory, (2) investigate characteristics of test items based on the results of the analysis using item response theory with the one parameter model (the Marginal Maximum Likelihood model), (3) compare the results of the item test analyses using classical test theory and item response theory, (4) investigate testees' ability, and (5) investigate equations for equating test sets for mathematics learning achievement for Grades VII, VIII, and IX of SMP Negeri 1 Tidore Kepulauan. The study was conducted using the quantitative method. The research population consisted of the students of SMP Negeri 1 Tidore Kepulauan with a total of 606 students. The parameter estimation was carried out on the basis of the parameter logistic models one and three using the BILOG 3.00 computer program. The data were processed and equated by means of the classical test theory approach (Microcat Iteman) and modern test theory. The results of the study have good quality. Based on classical test theory, the test set for Grade VII, VIII and IX consist of 29, 27 and 30 items respectively. Meanwhile, based on modern test theory, the test set for Grade VII, VIII and IX consists of 28, 27 and 24 items respectively. The equation for equating test sets from Grade VII to Grade VIII is $Q^ = 1.084X - 0.221$, that from Grade VIII to Grade IX is $R^* = 0.922X + 2.965$, and that from Grade VII to Grade IX is $R'^* = 0.998X + 2.775$. The testees ability was poor which had the average ability ranges from 0.265 to 0.58. This means that the testees' ability in the class examinations is still low. Based on the explanation of the test analyses using two approaches, namely classical test theory and modern test theory, it can be indicated that in the case of the mathematics subject for the three observed grades there is a difference in the percentages of good and valid items.*

Keywords: *vertical equating, learning achievement test, mathematics, item bank*

Pendahuluan

Pasal 66 (1) PP No. 19/2005 tentang Standar Nasional Pendidikan menyatakan bahwa penilaian hasil belajar oleh pemerintah bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan teknologi yang dilakukan dalam ujian nasional. Standar penilaian pendidikan merupakan standar nasional mengenai mekanisme, prosedur, dan instrumen dari penilaian hasil belajar siswa. Pada jenjang dasar dan menengah, penilaian hasil belajar dilakukan oleh guru, oleh satuan pendidikan dan oleh pemerintah. Penilaian hasil belajar guru dilakukan berkesinambungan untuk memantau proses, kemajuan, dan perbaikan hasil melalui ulangan harian, ulangan tengah semester, ulangan akhir semester, dan ulangan kenaikan kelas. Pencapaian kompetensi siswa, bahan penyusunan laporan pendidikan kemajuan hasil belajar, dan perbaikan pembelajaran digunakan sebagai dasar untuk melakukan penilaian.

Salah satu indikator yang menunjukkan masih rendahnya mutu pendidikan nasional adalah hasil penilaian internasional tentang prestasi siswa. Survei *Trends International Mathematics and Science Study* (TIMSS) tahun 2003 menempatkan Indonesia pada peringkat 34 dari 45 negara. Walau rerata skor naik menjadi 411 dari 403 pada tahun 1999, kenaikan ini secara statistik tidak signifikan dan skor itu di bawah rata-rata dari wilayah ASEAN. Prestasi itu bahkan relatif lebih buruk dalam *Programme for International Student Assessment* (PISA), yang mengukur kemampuan anak usia 15 tahun dalam hal literasi membaca, matematika, dan ilmu pengetahuan. Program yang dilaksanakan setiap tiga tahunan ini pada tahun 2003 menempatkan Indonesia pada peringkat 2 terendah dari 40 negara sampel, yaitu hanya satu peringkat lebih tinggi dari Tunisia (Kemendiknas, 2011, p.1).

Indonesia mengikuti TIMSS pada tahun 1999, 2003, 2007, 2011 dan PISA tahun 2000, 2003, 2006, 2009 dengan hasil tidak menunjukkan banyak perubahan pada setiap keikutsertaan. Pada PISA tahun 2009, Indonesia menduduki rangking 61 dari 65 peserta dengan rata-rata skor 371 dan rata-rata skor internasional 496. Prestasi pada TIMSS 2007 lebih memprihatinkan lagi, karena rata-rata skor siswa kelas 8 menurun menjadi 405 daripada tahun 2003, yaitu 411. Rangking

Indonesia pada TIMSS tahun 2007 mengenai penguasaan matematika untuk siswa sekolah menengah pertama menjadi urutan 36 dari 49 negara (Kemendiknas, 2011, p.1).

Tingginya kesenjangan mutu pendidikan tersebut menunjukkan rendahnya mutu pendidikan. Oleh karena itu, perlu ada perbaikan, baik di bidang manajemen pembelajaran, sistem pembelajaran, kualitas pendidik dan tenaga kependidikan, sarana dan prasarana, pendanaan maupun evaluasi pendidikan berkesinambungan. Penilaian merupakan komponen penting di dalam sistem pendidikan karena hasil penilaian mencerminkan perkembangan atau kemajuan hasil belajar pendidikan yang dapat dibandingkan dari waktu ke waktu, antara satu sekolah dengan sekolah yang lain, atau antara wilayah satu dengan wilayah yang lain. Proses penyetaraan tingkat pencapaian hasil pendidikan di antara sekolah atau wilayah ini dalam teori pengukuran disebut penyetaraan (*equating*).

Sekolah Menengah Pertama (SMP) Negeri 1 Tidore Kepulauan yang terletak di Kota Tidore Kepulauan Provinsi Maluku Utara dalam pelajaran matematika ditemukan memiliki persoalan terkait penyetaraan skor sebagai pengembangan sekaligus perbaikan pembelajaran bidang studi matematika secara internal sekolah sangat mendukung pencapaian standar nasional di bidang studi matematika. Persoalan yang muncul di mata pelajaran matematika adalah belum maksimalnya pemahaman guru terkait penyusunan kisi-kisi soal dan butir perangkat tes kepada siswa, Guru matematika dalam hal menciptakan perangkat tes yang standar untuk mengukur ciri atau kemampuan belajar siswa pada suatu jenjang tertentu belum sepenuhnya mampu melakukan hal tersebut, selanjutnya persoalan yang lain adalah kurangnya pemahaman guru matematika dalam melakukan pemilihan butir soal tes untuk peserta(siswa) tes tertentu agar tingkat kesukaraan butir pada perangkat tes itu cocok dengan tingkat kemampuan peserta (siswa) tes (perangkat tes yang adaptif).

Persoalan yang berikut rendahnya kemampuan guru matematika dalam membandingkan dua skor dari dua perangkat tes yang berbeda tetapi mengukur ciri atau kemampuan yang sama, dan selanjutnya sebagai persoalan berikut adalah rendahnya tingkat pemahaman guru matematika dalam membandingkan cara mendapatkan dua atau beberapa perangkat tes yang memiliki nilai

kesetaraan, sehingga siswa yang mengikuti perangkat tes yang berbeda tidak merasa dirugikan atau diuntungkan karena mengerjakan perangkat tes yang lebih sulit atau lebih mudah, serta belum maksimalnya pemahaman guru matematika dalam membedakan butir soal tes yang valid dan tidak valid.

Untuk itu diperlukan adanya penelusuran tingkat kemampuan tiap kelas suatu jenjang pendidikan, misalnya SMP, melalui penyetaraan vertikal. Umumnya penyusunan tes prestasi belajar, misalnya Ujian Nasional dan tes sumatif bersama, menyimpang dari teori pengukuran. Penyimpangan ini meliputi tiga hal: (1) tanpa penyetaraan, (2) tanpa ujicoba, dan (3) waktu yang tersedia untuk penulisan soal terlalu pendek. Karena itu, dalam melaksanakan evaluasi menggunakan beberapa perangkat tes, perlu dilakukan penyetaraan skor dari perangkat tes tersebut, karena dapat dijamin keadilan bagi peserta tes, dan kualitas proses belajar mengajar dapat diketahui (Crocker, 1986, Suryabrata, 1998).

Penilaian pembelajaran diklasifikasi menjadi dua, yaitu penilaian formatif dan penilaian sumatif. Penilaian formatif bertujuan memperbaiki proses belajar dan mengajar. Hasil penilaian formatif dianalisis untuk mengetahui konsep yang belum dipahami sebagian siswa dan diikuti kegiatan remedial. Kegiatan remedial merupakan kegiatan pembelajaran untuk mengatasi kesulitan belajar siswa yang diidentifikasi dari kegiatan penilaian formatif. Penilaian sumatif bertujuan untuk menetapkan tingkat keberhasilan siswa.

Tes adalah teknik atau cara yang digunakan dalam pelaksanaan kegiatan pengukuran, termasuk berbagai pertanyaan, pernyataan, atau rangkaian tugas yang harus dikerjakan dan dijawab oleh siswa untuk mengukur aspek perilaku. Dalam rumusan ini, terdapat beberapa unsur: (1) tes adalah cara atau teknik yang disusun secara sistematis dan digunakan dalam pengukuran, (2) di dalam tes, ada berbagai pertanyaan, pernyataan atau serangkaian tugas yang harus dijawab dan dikerjakan siswa, (3) tes digunakan untuk mengukur aspek perilaku siswa, dan (4) hasil tes siswa perlu diberi skor dan nilai.

Tes berkualitas berkaitan dengan soal berkualitas. Skor yang berkualitas memenuhi kriteria sebagai alat ukur yang baik. Kriteria alat ukur yang baik dapat ditinjau dari teori tes klasik (*classical tes theory—CTT*) dan teori tes modern atau teori respon butir (*item response*

theory—IRT). Pada prinsipnya, alat ukur harus mempunyai bukti kesahihan dan keandalan. Kesahihan alat ukur dapat dilihat dari konstruk alat ukur, sejauhmana dapat mengukur seperti apa yang direncanakan.

Parameter yang diukur pada CTT adalah reliabilitas tes, indeks daya beda butir (*discrimination index*), index kesukaran butir (*difficulty index*), validitas isi (*content validity*), validitas konstruk (*construct validity*), dan validitas kriteria (*criterion related validity*). Parameter yang diukur pada IRT adalah kemampuan (*ability*) dan parameter yang terdiri dari indeks daya beda butir (*discrimination index*), indeks kesukaran butir (*diffulty indeks*), dan terkaan (*guessing*). Salah satu kriteria alat ukur yang baik adalah memiliki terkaan (*guessing*) yang relatif kecil.

Teori tes klasik digunakan secara luas dalam pengukuran pendidikan di Indonesia. Salah satu keterbatasan teori ini adalah ketika tes yang sama diberikan kepada siswa yang berbeda dan hasilnya dinyatakan dengan skor total. Tingkat kemampuan siswa tidak bisa dibandingkan berdasarkan skor total yang diperoleh siswa, karena skor total tidak menunjukkan tingkat kesulitan tes yang dikerjakan. Siswa yang kemampuannya lebih rendah mungkin mendapatkan skor lebih tinggi dalam tes yang mudah dibandingkan siswa lain yang kemampuannya lebih tinggi tetapi mendapatkan tes yang lebih sulit. Keadaan ini menggambarkan bahwa skor total tidak dapat dibandingkan karena dua siswa tersebut mengerjakan butir yang mungkin memiliki tingkat kesulitan berbeda, walau skor total yang dicapai sama.

Dalam kegiatan pembelajaran di kelas, guru menggunakan teori tes klasik untuk mengidentifikasi karakteristik soal yang digunakan, serta mengidentifikasi tingkat kemampuan siswa berdasarkan skor total yang diperoleh siswa. Pada IRT, siswa mempunyai sifat invarian terhadap tingkat kesulitan butir, dan sebaliknya. Tingkat kemampuan siswa tidak tergantung tingkat kesulitan butir, dan tingkat kesulitan butir tidak tergantung pada siswa yang mengerjakannya.

Penggunaan skor dari dua tes yang disetarakan menuntut keduanya harus memiliki tingkat kesulitan setara. Proses statistik yang biasanya digunakan untuk menyetarakan skor kedua tes tersebut disebut penyetaraan (*equating*). Hambleton dan Swaminathan (1985, p.23) menyatakan bahwa penyetaraan

adalah prosedur statistik untuk menetapkan hubungan di antara skor dari dua tes atau lebih. Kolen dan Brenan (1995, p. 2) menyatakan bahwa proses penyetaraan bisa dilakukan untuk menyetarakan dua tes atau lebih dengan materi dan tingkat kesulitan yang setara.

Pada penggunaan skala kemampuan yang sama pada penyetaraan skor tes, ada keuntungan, misalnya, memungkinkan dilakukan evaluasi atas hasil tes, dapat mengembangkan tes, dan dapat dikembangkannya bank soal. Menurut Hambleton dan Swaminathan (1985, p.20), prosedur penyetaraan skor tes (*test equating*) dapat dilakukan dengan dua cara, yaitu penyetaraan vertikal dan penyetaraan horisontal. Penyetaraan vertikal dilakukan untuk menyetarakan skor tes dalam dua tes atau lebih yang tingkat kesulitannya dirancang berbeda, tetapi mengukur isi dan jenis kemampuan yang sama. Penyetaraan ini dirancang untuk melihat kontinuitas tes. Kontinuitas tes mengacu pada keberlanjutan tes yang digunakan untuk mengukur perkembangan atau perubahan tingkat kemampuan siswa. Penyetaraan horisontal dilakukan pada tes paralel yang memiliki kesamaan isi dan tingkat kesulitan, dan diberikan kepada kelompok siswa yang memiliki tingkat kemampuan setara.

Menurut Kolen dan Brenan (1995, p.13) untuk mengurangi ketidakakuratan hasil penyetaraan, diperlukan rancangan penyetaraan. Rancangan penyetaraan itu beragam, yaitu grup tunggal (*Single Grup Design—SG*), rancangan grup ekuivalen (*Equivalent Group Design—EG*), rancangan grup seimbang (*Counter Balanced Design—CB*), rancangan pengait (*Anchor Test Design—AT*) atau (*Non-Equivalent Anchor Test Design—NEAT*). Rancangan tersebut memiliki karakteristik berbeda serta memiliki kelebihan dan kekurangan sendiri.

Penyetaraan skor tes menggunakan IRT adalah model penyetaraan yang lebih representatif dibandingkan model penyetaraan dengan CTT. IRT memiliki sifat invarian pada parameter. Parameter kemampuan (*ability*) siswa bersifat invarian parameter tes, dan sebaliknya. Oleh karena itu, tes yang dikerjakan siswa tetap pada skala yang sama selama fungsi informasi tes tinggi. Berkaitan dengan hal tersebut, maka tujuan penelitian ini adalah menemukan keakuratan estimasi parameter butir pada tes penyetaraan menggunakan metode kurva karakteristik butir (*Item Charac-*

teristic Curve—ICC) yang ditunjukkan oleh RMSD parameter butir sebelum dan sesudah penyetaraan. Selain itu, penelitian ini juga bertujuan untuk menemukan sensitivitas metode linear yang terdiri atas *Tucker-Levine score method* dan *Levine true score method* dan metode *equipercentile equating method* yang ditunjukkan oleh nilai RMSE sebelum dan setelah penyetaraan.

Penelitian ini menekankan perbaikan mutu pendidikan yang difokuskan pada penyusunan soal tes kemampuan matematika berkualitas untuk mengetahui peningkatan kemampuan siswa melalui perbaikan sistem pengujian, yaitu dengan menyediakan soal tes kemampuan matematika. Pengembangan sistem pengujian dengan tes dipilih karena tes bisa memantau perkembangan siswa dalam berbagai tahapan. Karena itu, diperlukan strategi pengembangan tes agar diperoleh manfaat optimal bagi upaya peningkatan mutu pendidikan.

Permasalahan peningkatan mutu pendidikan khususnya kualitas SMP, perlunya standarisasi skor batas kelulusan atau kenaikan kelas dipandang penting untuk melakukan penelitian tentang penyetaraan (*equating*), sementara itu pentingnya memilih metode yang tepat dalam menentukan skor batas kelulusan atau kenaikan kelas menuntut perlunya dilakukan komparasi dari berbagai metode. Dari permasalahan di atas maka dilakukan penelitian tentang “Komparasi Metode Penyetaraan Vertikal Skor pada Mata Pelajaran Matematika Siswa SMP Negeri 1 Tidore Kepulauan”.

Metode Penelitian

Jenis Penelitian

Penelitian ini dilakukan dengan metode kuantitatif. Metode ini digunakan untuk membandingkan skor siswa dari beberapa perangkat tes matematika dengan menggunakan metode penyetaraan vertikal skor pada SMP Negeri 1 Tidore Kepulauan Tahun Pelajaran 2012/2013.

Waktu dan Tempat Penelitian

Penelitian ini akan dilaksanakan di SMP Negeri 1 Kota Tidore Kepulauan Tahun Pelajaran 2012/2013, semester ganjil. Lama penelitian bulan Maret-April 2013. Populasi dan Sampel Penelitian

Populasi penelitian ini adalah siswa SMP Negeri 1 Tidore Kepulauan, dengan jumlah total sebanyak 606 siswa. Sampel dari populasi ini diambil dengan teknik acak proporsional bertingkat (*stratified proporsional random sampling*). Sampel penelitian ini adalah siswa kelas VII, VIII, dan IX, masing-masing 2 kelas. Sampel siswa di setiap kelas diambil secara bergantian, yaitu 32 siswa dari kelas VII untuk perangkat tes X, 30 siswa kelas VIII untuk perangkat tes Y, dan 27 siswa kelas IX untuk perangkat tes Z.

Teknik Pengumpulan dan Teknik Analisis Data

Sumber data dalam penelitian ini berupa lembar jawaban tes kemampuan matematika yang memenuhi prinsip-prinsip penyetaraan. Jenis data berupa data primer yang bersifat kuantitatif.

Penelitian ini menggunakan alat pengumpul data dengan teknik dokumentasi. Data penelitian yang dianalisis berupa paket soal yang terdiri dari: (1) kisi-kisi dan lembar soal, (2) respons siswa pada mata pelajaran matematika.

Instrumen yang digunakan untuk mengambil data yaitu berupa tes hasil belajar yang dapat mengukur kemampuan penalaran siswa. Tipe tes yang dikembangkan adalah tes objektif berbentuk pilihan ganda dengan 4 *option*. Soal bentuk pilihan ganda terdiri dari batang tubuh soal (*stem*), yang berupa pertanyaan pengantar atau pernyataan tak lengkap dan dua lebih kemungkinan jawaban. Secara teknis kemungkinan jawaban yang benar disebut dengan kunci jawaban, dan yang lainnya disebut pengecoh (*distractor*). Kesemua kemungkinan jawaban yang disebut *option*.

Instrumen penelitian ini dikembangkan dalam beberapa tahap sebagai berikut: 1) penulisan kisi-kisi soal, 2) telaah butir soal, 3) ujicoba Instrumendan 4) analisis butir soal

Analisis data dalam penelitian ini dilakukan dengan pendekatan teori tes klasik dengan menggunakan bantuan program MicroCat ITEMAN dan teori tes modern dengan menggunakan bantuan program BILOG 3.00.

Variabel Penelitian

Variabel yang diukur dalam penelitian ini adalah a) karakteristik butir soal, meliputi: tingkat kesukaraan, daya pembeda, b)

skala penyetaraan vertikal, meliputi parameter ciri peserta/kemampuandanciri butir/tingkat kesukaraan dan 3) perkembangan prestasi belajar: diperoleh dari skor hasil tes siswa terhadap *anchor*.

Perangkat tes mata pelajaran matematika dalam penelitian ini terdiri dari 3 paket tes pilihan ganda yang dikembangkan oleh peneliti bekerja sama dengan guru mata pelajaran akan validasi. Perangkat tes Matematika kelas VII terdiri dari 30 butir soal dengan *anchor* sebanyak 12 butir, perangkat tes kelas VIII terdiri 30 butir dengan *anchor* sebanyak 9 butir, dan perangkat tes kelas IX terdiri 30 butir dengan *anchor* sebanyak 4 butir.

Definisi Operasional Variabel Penelitian

Untuk memperjelas pembahasan dalam penelitian ini, maka diberikan beberapa definisi operasional adalah: a) karakteristik butir soal adalah proporsi peserta yang menjawab benar, b) penyetaraan vertikal (*vertical equating*) adalah suatu prosedur empirik dalam rangka menyetarakan skala kemampuan siswadari kelas yang satu ke kelas yang lain, dan tingkat kesukaraan butir dari perangkat tes satu ke perangkat yang lain untuk berbagai tingkatandanc) tes prestasi belajar adalah suatu bentuk tes yang digunakan untuk mendapatkan data sebagai bahan informasi tentang seberapa banyak pengetahuan yang dimiliki dan dikuasai oleh seseorang sebagai akibat dari program pendidikan.

Hasil Penelitian dan Pembahasan

Hasil analisis terhadap uji coba soal Matematika Kelas VII menunjukkan hasil analisis terhadap perangkat tes Matematika kelas VII sebagai berikut: butir yang termasuk kategori mudah sebanyak 1 butir; sedang sebanyak 22 butir dan sukar sebanyak 7 butir. Butir soal yang memiliki daya pembeda yang memenuhi kriteria sebanyak 22 butir dan yang revisi sebanyak 8 butir.

Hasil analisis terhadap uji coba soal Matematika Kelas VIII menunjukkan hasil analisis terhadap perangkat tes Matematika kelas VIII sebagai berikut: butir yang termasuk kategori mudah sebanyak 0 butir; sedang sebanyak 26 butir dan sukar sebanyak 4 butir. Butir soal yang memiliki daya pembeda yang memenuhi kriteria sebanyak 24 butir dan yang revisi sebanyak 6 butir

Hasil analisis terhadap uji coba soal Matematika Kelas Ix menunjukkan hasil analisis terhadap perangkat tes Matematika kelas IX sebagai berikut: butir yang termasuk kategori mudah sebanyak 1 butir; sedang sebanyak 19 butir dan sukar sebanyak 10 butir. Butir soal yang memiliki daya pembeda yang memenuhi kriteria sebanyak 28 butir dan yang revisi sebanyak 2 butir.

Tabel 1. Hasil Analisis Teori Tes Klasik dengan ITEMAN versi 3.0

Kelas	Mata Pelajaran	Jumlah Peserta	Jumlah Butir	Nilai Ststistik Alpha	SEM	M _p	M _{bis}
VII	Matematika	64	30	0,852	2,400	0,563	0,390
VIII	Matematika	60	30	0,856	2,403	0,565	0,413
IX	Matematika	54	30	0,843	2,322	0,562	0,419

Dari Tabel 1 terlihat nilai M_{bis} untuk mata pelajaran Matematika kelas VII, VIII, dan IX tahun pelajaran 2012/2013 ujian semester ganjil sama dengan atau lebih dari 0,390. Nilai M_{bis} adalah rata daya bedadari koefisien korelasi biserial. Daya beda yang dianggap memadai untuk suatu butir soal adalah daya beda yang sama atau lebih besar dari 0,390.

Berdasarkan hasil analisis ITEMAN terhadap butir soal mata pelajaran Matematika dapat dikatakan bahwa secara umum seluruh butir soal yang diujikan pada ujian tahun pelajaran 2012/2013 ujian semester ganjil memiliki daya beda yang sangat memadai. Hal iniberarti butir-butir soal yang diujikan memiliki kemampuan membedakan kelompok peserta tes yang berprestasi tinggi dengan kelompok peserta tes yang berprestasi rendah. Koefisien indeks daya beda pada semua kelas memiliki nilai positif. Artinya peserta ujian yang menjawab benar butir soal mempunyai korelatif tinggi dalam ujian tersebut.

Dilihat dari tingkat kesukarannya, rata-rata tingkat kesukaran butir soal-soal ujian dari mata pelajaran Matematika pada tahun pelajaran 2012/2013 ujian semester ganjil cenderung mudah. Hal ini ditunjukkan oleh nilai koefisien M_p yaitu tingkat kesukaran butir soal yang 1 memiliki nilai antara 0,562 sampai dengan 0,565 (M_p > 0,5).

Dari ketiga kelas, rata-rata butir soal yang memiliki tingkat kesukaran paling rendah yaitu pada kelas VIII memiliki tingkat kesukaran 0,562.

Hasil reliabilitas soal kelas VII, VIII dan IX diketahui nilai kofisien reliabilitas antara 0,843 sampai dengan 0,856 (p > 0,70) sehingga dinyatakan reliabel.

Berdasarkan nilai daya beda, tingkat kesukaran dan reliabilitas tersebut dapat diketahui soal-soal yang valid danyang tidak valid. Butir-butir soal mata pelajaran Matematika baik adalah pada semua kelas VII, VIII dan IX dengan persentase butir-butir soal yang valid di atas 80,0%. Sementara itu, mata pelajaran Matematika kelas IX memberikan 93,33% butir-butir soal yang valid. Kriteria soal yang valid dalam hal ini dilihat dari nilai biserialnya (M_{bis}). Dalam hal ini, butir soal dinyatakan valid jika nilai M_{bis} ≥ 0,2.

Hasil Analisis Tes Modern (Marginal Maximum Likelihood) terhadap Butir soal Mata Pelajaran Matematika

Dalam penelitian ini, tes modern yang digunakan adalah *Marginal Maximum Likelihood*. Perangkat lunak yang digunakan untuk analisis butir soal dengan *Marginal Maximum Likelihood* adalah model BILOG versi 3.0.

Perangkat Tes Matematika Kelas VII

Analisis Butir berdasarkan nilai-nilai parameter

Analisis butir ini dilakukan untuk memilih butir-butir yang akan dilibatkan dalam analisis berikutnya yaitu penyetaraan skor, yang berdasarkan kriteria sebagai berikut: 1) memiliki parameter “a” antara 0 sampai 2, atau 0 < a < 2; 2) memiliki parameter “b” antara -2 sampai 2, atau -2 ≤ b ≤ 2; 3) memiliki parameter “c” antara 0 sampai 1, atau 0 < c < 1; nilai X_{hitung} > X_{tabel} pada taraf signifikansi 0,01. Berdasarkan ketiga kriteria tersebut, bila suatu butir tidak memenuhi salah satu atau lebih dari kriteria yang telah ditetapkan, maka butir tersebut tidak dilibatkan dalam penyetaraan skor.

Rangkuman hasil estimasi soal Kelas VII yang terdiri dari daya pembeda (a) dan tingkat kesukaran (b) soal untuk model logistik satu parameter dan kecocokan model. Semua butir- butir pada perangkat kelas VII memenuhi kriteria untuk dilanjutkan ke analisis berikutnya. Pada perangkat tes Matematika kelas VII semua butir akan diikuti dalam analisis berikutnya (penyetaraan)

sehingga untuk perangkat kelas VII yang dapat dilanjutkan ke analisis berikutnya 23 butir.

Daya beda terendah adalah dengan nilai 0,673 dan yang tertinggi dengan nilai 0,673. Tingkat kesukaran tertinggi sebesar 1,112 dimiliki oleh butir nomor 6, 28 dan tingkat kesukaran terendah dimiliki butir 10 dan 17 dengan nilai 1,107.

Perangkat Tes Matematika Kelas VIII

Analisis butir berdasarkan nilai-nilai parameter

Rangkuman hasil estimasi soal Kelas VIII yang terdiri dari daya pembeda (a), tingkat kesukaran (b), pendugaan semu (c) soal untuk model logistik tiga parameter dan kecocokan model. Pada perangkat tes Matematika kelas VIII ada 3 butir yang tidak memenuhi kriteria untuk dilanjutkan ke analisis berikutnya. Butir-butir tersebut adalah sebagai berikut: butir 5 dengan $a = 2.392$, butir 15 dengan $a = 2.265$ dan butir 17 dengan $a = 7.456$. Sehingga untuk perangkat kelas VIII yang dapat dilanjutkan ke analisis berikutnya 21 butir.

Daya beda terendah adalah dengan nilai 0,674 dan yang tertinggi dengan nilai 7,456. Tingkat kesukaran tertinggi sebesar 1,564 dimiliki oleh butir nomor 23 dan tingkat kesukaran terendah dimiliki butir 10 dengan nilai -0,036.

Perangkat Tes Matematika Kelas IX

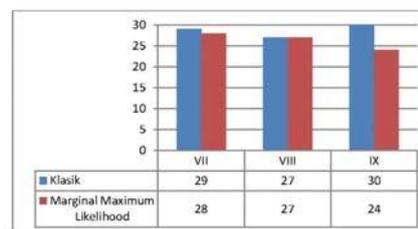
Analisis Butir berdasarkan nilai-nilai parameter

Rangkuman hasil estimasi soal Kelas IX yang terdiri dari daya pembeda (a) dan tingkat kesukaran (b) soal untuk model logistik satu parameter dan kecocokan model. Pada perangkat tes Matematika kelas IX ada 1 butir yang tidak memenuhi kriteria untuk dilanjutkan ke analisis berikutnya. Butir-butir tersebut adalah sebagai berikut: butir 26 dengan $b = 2,215$. Sehingga untuk perangkat kelas IX yang dapat dilanjutkan ke analisis berikutnya 26 butir.

Daya beda terendah dan tertinggi adalah dengan nilai 0,604. Tingkat kesukaran tertinggi sebesar 2,215 dimiliki oleh butir nomor 26 dan tingkat kesukaran terendah dimiliki butir 10 dengan nilai -1,146.

Perbandingan Hasil Analisis Tes Klasik dan Tes Modern (Marginal Maximum Likelihood) terhadap Butir Soal Matematika.

Gambaran jumlah butir soal yang valid dengan jumlah peserta ujian tahun pelajaran 2012/2013 ujian semester berdasarkan model klasik dan modern.



Gambar 1. Perbandingan Jumlah Butir Soal yang Valid Model Klasik dan Marginal Maximum Likelihood ujian Matematika tahun pelajaran 2012/2013 ujian semester ganjil

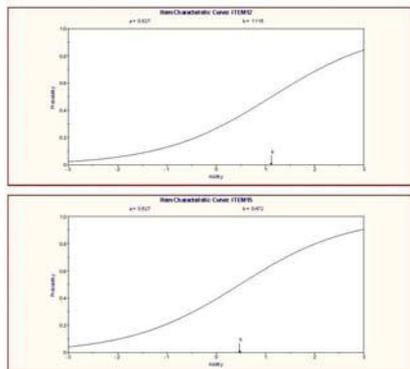
Dari Gambar 1. terlihat untuk mata pelajaran Matematika kelas VII jumlah pesertanya 64 siswa, kelas VIII jumlah peserta 60 siswa, dan kelas IX jumlah peserta 54 siswa. Persentase valid pada mata pelajaran Matematika kelas VII sebanyak 93,33%, pada mata pelajaran Matematika kelas VIII sebanyak 90,00% dan mata pelajaran matematika kelas IX sebanyak 80%. Hal ini menunjukkan bahwa jumlah peserta ujian berpengaruh terhadap hasil analisis kedua model tersebut, terutama pada *Model Marginal Maximum Likelihood*. Secara umum dapat dilihat bahwa pada jumlah peserta ujian yang sama dengan 60 dengan model *Marginal Maximum Likelihood* memberikan analisis butir sama dengan baik. Sementara itu, pada peserta ujian yang kurang dari 60 analisis butir soal dengan menggunakan model klasik memberikan analisis yang lebih baik.

Karakteristik Soal dan Estimasi Kemampuan Peserta Ujian dengan Menggunakan Marginal Maximum Likelihood

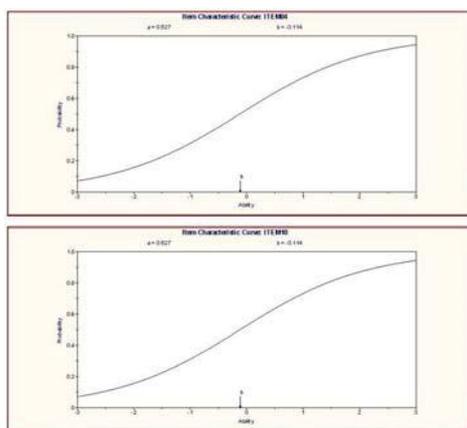
Dengan menggunakan analisis model *Marginal Maximum Likelihood*, pada setiap butir soal dapat diketahui karakteristik soal dan estimasi kemampuan peserta ujian. Secara rinci, karakteristik butir soal mata pelajaran Matematika tahun pelajaran 2012/2013 ujian semester ganjil terlihat pada lampiran 2. Seperti yang telah dipaparkan pada uraian sebelumnya, bahwa item respon teori (dalam penelitian ini model *Marginal Maximum Likelihood* sangat baik digunakan untuk jumlah peserta ujian yang banyak. Berdasarkan mata pelajaran Matematika tahun pelajaran 2012/2013 ujian semester

ganjil peserta ujian terbanyak adalah kelas VII dan kelas VIII.

Analisis dengan menggunakan BILOG 3.00 dapat menghasilkan karakteristik butir soal dan kemampuan peserta ujian (θ). Karakteristik butir soal dapat digambarkan melalui kurva karakteristik butir. Soal yang baik memiliki bentuk huruf S sebagaimana terlihat pada gambar.



Gambar 2. Kurva karakteristik Butir Soal yang Baik Mata Pelajaran Matematika Kelas VII tahun pelajaran 2012/2013



Gambar 3. Kurva karakteristik Butir Soal yang Baik Mata Pelajaran Matematika Kelas VII 2012/2013 ujian semester ganjil

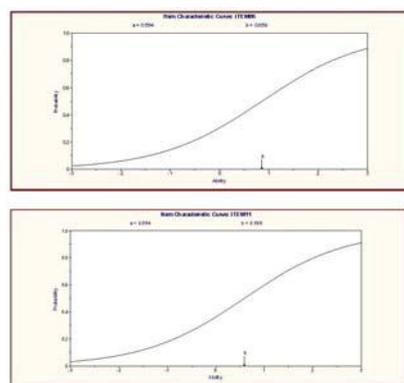
Dari gambar 2 dapat dijelaskan bahwa pada ujian matematika siswa kelas VII dengan kemampuan $-0,5$ maka peluang menjawab benar soal no.12 sekitar 0,2. Artinya hanya 20% siswa dengan kemampuan $-0,5$ dapat menjawab dengan benar soal no. 12 tersebut.

Demikian juga halnya dengan kemampuan siswa $0,5$ maka peluang menjawab benar soal tersebut sekitar 3,9. Artinya hanya sekitar 39% siswa yang menjawab benar soal no. 12 dengan kemampuan $0,5$. Sementara itu, pada

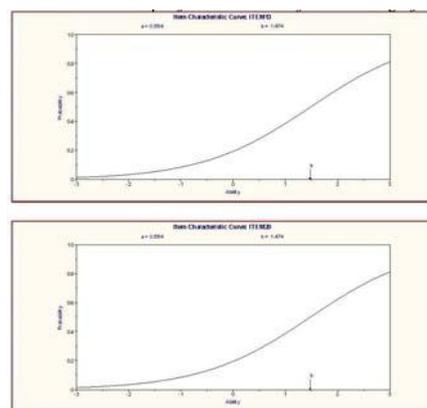
soal no. 15, peluang menjawab dengan benar siswa dengan kemampuan $-0,5$ sebesar 0,24 (sebanyak 24%), sedangkan peluang menjawab benar siswa dengan kemampuan $0,5$ sebesar 0,45 (sebanyak 45%).

Hal ini menunjukkan bahwa soal no 12 lebih sulit dibandingkan soal no. 15. Lain halnya dengan soal pada 6terlihat bahwa tingkat kesukaran kedua soal tersebut cenderung sama. Hal ini dapat dilihat dari peluang menjawab benar pada tingkat kemampuan tertentu relatif sama.

Karakteristik butir soal yang baik mata pelajaran Matematika kelas VIII disajikan pada kurva sebagai berikut.



Gambar 4. Kurva karakteristik Butir Soal yang Baik Mata Pelajaran Matematika Kelas VIII 2012/2013 ujian semester ganjil



Gambar 5. Kurva karakteristik Butir Soal yang Baik Mata Pelajaran Matematika Kelas VIII 2012/2013 ujian semester ganjil

Dari gambar 4 dapat dijelaskan bahwa pada ujian matematika siswa kelas VII dengan kemampuan $-0,5$ maka peluang menjawab benar soal no 5 sekitar 0,23. Artinya hanya

23% siswa dengan kemampuan $-0,5$ dapat menjawab dengan benar soal no 5. Demikian juga halnya dengan kemampuan siswa $0,5$ maka peluang menjawab benar soal tersebut sekitar $0,41$. Artinya hanya sekitar 41% yang menjawab benar soal no 5 dengan kemampuan $0,5$. Sementara itu, pada soal no 11, peluang menjawab dengan benar siswa dengan kemampuan $-0,5$ sebesar $0,28$ (sebanyak 28%), sedangkan peluang menjawab dengan benar siswa dengan kemampuan $0,5$ sebesar $0,49$ (sebanyak 49%). Hal ini menunjukkan bahwa soal no 5 lebih sulit dibandingkan dengan soal no 11. Lain halnya dengan soal pada 8 terlihat bahwa tingkat kesukaran kedua soal tersebut cenderung sama. Hal ini dapat dilihat dari peluang menjawab benar pada tingkat kemampuan tertentu raltif sama.

Tabel 2. Statistik Deskriptif Kemampuan Siswa Kelas VII, VIII dan IX (θ)

Klas	N	Min	Max	Mean	Std. Deviation
VII	30	-0,87	1,32	0,5834	0,54465
VIII	30	-1,26	1,47	0,2650	0,69686
IX	30	-1,21	2,34	0,4467	1,03648
Valid N (listwise)	30				

Penyetaraan Perangkat Tes

Proses penyetaraan dapat digunakan untuk mengukur perkembangan dan pemetaan berbagai kecenderungan serta mendapat dan menggabungkan informasi dari perangkat tes berbeda. Penyetaraan menurut Linn (1989) merupakan metode empiris yang diperlukan untuk mentransformasikan skor tes ke skor tes yang lain. Kegunaan hasil penyetaraan dapat digunakan untuk membandingkan atau konversi antara dua kemampuan yang mengukur perangkat tes.

Pada penyetaraan tingkat kesukaran antartest dan distribusi dapat dibedakan. Hal ini sesuai dengan pernyataan bahwa tes A dan tes B yang non paralel dapat disetarakan jika tes tersebut mengukur kemampuan yang sama. Cara memperoleh skor yang sama, kemampuan skor tes A dan tes B dikonversikan pada skala utama (*common score scale*).

Perangkat tes yang digunakan dalam penelitian ini yaitu rancangan A tipe I. Tes tipe I dibentuk oleh dua perangkat tes, yakni perangkat tes kelas VII (P) dan VIII (Q), kelas

VIII dan IX (R), serta kelas VII dan IX (R'). Skor peserta tes pada perangkat tes dinyatakan dalam skala 0-22.

Tabel 3. Penyetaraan Berupa Rancangan A

Jenis Rancangan	Kelompok	
	K1	K2
A	X	Y

Rumus konversi skor dari kelas VII (P) ke kelas VIII (Q), perangkat tes VII sebagai pembanding, yaitu $Q^* = a(P-c) + d$. Dengan a , c , dan d sebagai berikut:

$$a = \frac{\sigma q}{\sigma p}$$

$$c = \mu p$$

$$d = \mu q$$

Konversi skor dari kelas VII (Q) ke kelas IV (R), perangkat tes VIII sebagai pembanding, yaitu $R^* = a(Q-c) + d$. Dengan a , c , dan d sebagai berikut:

$$a = \frac{\sigma R}{\sigma Q}$$

$$c = \mu Q$$

$$d = \mu R$$

Konversi skor dari kelas VII ke kelas IX (perangkat tes VII sebagai pembanding), yaitu $R'^* = a(P-c) + d$. Dengan a , c , dan d sebagai berikut:

$$a = \frac{\sigma R}{\sigma P}$$

$$c = \mu p$$

$$d = \mu R$$

Pada penyetaraan P ke Q diperoleh nilai a , c , dan d masing-masing sebesar $a = 1,084$, $c = 8,891$, $d = 9,417$. Angka-angka tersebut dimasukkan ke dalam persamaan $Q^* = a(P-c) + d$, sehingga diperoleh persamaan konversi tes mata pelajaran Matematika siswa-siswi dari Sekolah Menengah Pertama Negeri 1 Tidore Kepulauan dari perangkat tes P ke Q, yaitu $Q^* = 1,084(P - 8,891) + d$.

Memperhatikan persamaan regresi yang diperoleh, skor yang diperoleh seseorang peserta tes dengan mengikuti perangkat tes P dan skor peserta tes yang mengikuti perangkat tes Q terdapat perbedaan. Misalnya skor pada perangkat tes P = 5, sama kedudukannya dengan skor pada perangkat Q = 5,2.; skor pada perangkat P = 6 sama kedudukannya dengan skor pada perangkat Q = 6,283.

Sebaliknya dengan cara yang sama dapat ditentukan pula persamaan konversi skor

dari perangkat tes Q ke perangkat tes P (Q sebagai pembanding, yaitu $P^* = 0,992 (Q - 9,417) + 8,891$ atau $P^* = 1,084X + 0,526$. Persamaan ini sama dengan persamaan yang diperoleh dari invers persamaan konversi skor penyetaraan perangkat tes P ke Q ($Q^* = a(P - c) + d$). Misalnya, skor pada Q = 5,2, akan sama dengan skor pada P = 5. Perbedaan ini pada kedua model jika saling dikonversikan pada kedua persamaan skor penyetaraan hanya berbeda dalam pembulatan angka desimal.

Penyetaraan Q ke R diperoleh nilai a, c, dan d masing-masing sebesar a = 0,922, c = 9,417, d = 11,648. Angka-angka tersebut dimasukkan ke dalam persamaan $R^* = a(Q - c) + d$, sehingga diperoleh persamaan konversi tes mata pelajaran Matematika siswasiswi dari Sekolah Menengah Pertama Negeri 1 Tidore Kepulauan dari perangkat tes Q ke R, yaitu $R^* = 0,922 (P - 9,417) + 11,648$ atau $R^* = 0,922X + 10,7063$.

Memperhatikan persamaan regresi yang diperoleh, skor yang diperoleh seseorang peserta tes dengan mengikuti perangkat tes Q dan skor peserta tes yang mengikuti perangkat tes R terdapat perbedaan. Misalnya skor pada perangkat tes Q = 10, sama kedudukannya dengan skor pada perangkat R = 12,185. Sebaliknya dengan cara yang sama dapat ditentukan pula persamaan konversi skor dari perangkat tes R ke perangkat tes Q (R sebagai pembanding, yaitu $Q^* = 1,085 (Q - 11,648) + 9,417$. Persamaan ini sama dengan persamaan yang diperoleh dari invers persamaan konversi skor penyetaraan perangkat tes Q ke R ($R^* = a(Q - c) + d$).

Pada penyetaraan P ke R' diperoleh nilai a, c, dan d masing-masing sebesar a = 0,998, c = 8,891, d = 11,648. Angka-angka tersebut dimasukkan ke dalam persamaan $R'^* = a(P - c) + d$, sehingga diperoleh persamaan konversi tes mata pelajaran Matematika siswasiswi dari Sekolah Menengah Pertama Negeri 1 Tidore Kepulauan dari perangkat tes P ke R', yaitu $R'^* = 0,998 (P - 8,891) + d$ atau 11,648 atau $R'^* = 0,998X + 2,757$.

Berdasarkan persamaan regresi yang diperoleh, skor yang diperoleh seseorang peserta tes dengan mengikuti perangkat tes P dan skor peserta tes yang mengikuti perangkat tes R' terdapat perbedaan. Misalnya skor pada perangkat tes P = 9, sama kedudukannya dengan skor pada perangkat Q = 11,757; skor pada perangkat P = 6 sama kedudukannya dengan skor pada perangkat Q = 8,761.

Sebaliknya dengan cara yang sama dapat ditentukan pula persamaan konversi skor dari perangkat tes R' ke perangkat tes P (R' sebagai pembanding, yaitu $P^* = 1,001 (Q - 11,648) + 8,891$. Persamaan ini sama dengan persamaan yang diperoleh dari invers persamaan konversi skor penyetaraan perangkat tes P ke R' ($R'^* = a(P - c) + d$). Misalnya, skor pada Q = 11,757, akan sama dengan skor pada P = 9. Apabila terdapat perbedaan ini pada kedua tidak terlalu jauh atau hanya dalam pembulatan angka desimal.

Dalam penyetaraan perangkat tes secara empiris telah ditunjukkan pasangan skor antara perangkat tes P ke perangkat tes Q, perangkat tes Q ke perangkat tes R, dan P ke perangkat tes R'. Persamaan konversi yang diperoleh menggambarkan bahwa skor P dan Q, Q dan R, setara P dan R memiliki perbedaan. Misalnya untuk P = 20, skor pada Q = 21,458. Demikian juga sebaliknya untuk Q = 10, skor pada P = 11,445. Perbedaan tersebut artinya, peserta tes yang menggunakan perangkat tes P (soal Matematika kelas VII) skor yang diperoleh 20 dikonversikan perangkat tes Q (yang digunakan di kelas VIII) setara dengan skor 21,458. Sebaliknya, peserta tes yang menggunakan perangkat tes Q (peserta tes dikelas VIII), skor yang diperoleh 10 dikonversikan kepada perangkat tes P (yang digunakan kelas VII) setara dengan 11,445.

Perbedaan skor dari hasil konversi Q^* dan P^* , dapat diartikan juga bahwa kedua perangkat tes tersebut memiliki tingkat kesukaran yang berbeda pula. Bila dilihat dari indeks kesukaran tes antara perangkat tes P, Q, dan R masing-masing 0,563; 0,565; dan 0,562. Sehingga sangat wajar rerata skor berbeda namun perbedaan yang tidak terlalu jauh antara perangkat tes P (peserta kelas VII), rerata skor perangkat tes Q (peserta kelas VIII), dan rerata skor perangkat tes R (peserta kelas VIII).

Besarnya nilai rerata kesalahan pengukuran penyetaraan pada persamaan konversi hasil penyetaraan perangkat tes mata pelajaran Matematika Sekolah Menengah Pertama Negeri 1 Tidore Kepulauan yaitu masing-masing sebesar $SEE[Q(P_0)] = 3,432$ dan $SEE[Q(P_0)] = 3,458$; $SEE[R(Q_0)] = 3,186$ dan $SEE[R(Q_0)] = 4,496$; dan $SEE[P(R_0)] = 3,100$ dan $SEE[P(R_0)] = 4,148$.

Berdasarkan temuan penelitian dari hasil analisis data yang telah disebutkan diatas

dapat dikemukakan penjelasan dan pembahasan seperti berikut:

Karakteristik butir soal

Teori tes klasik menyatakan bahwa karakteristik butir soal dapat dilihat dari tingkat kesukaraan butir, daya beda, keberfungsian pengecoh dan reliabilitas tes. Menurut teori tes modern, karakteristik butir soal dapat dilihat dari peluang menjawab benar, tingkat kesukaraan butir (b), daya beda (a), dan tebakan (c). Hasil analisis berdasarkan teori tes modern, karena cara perolehannya berbeda. Dengan menggunakan teori tes modern, karena cara perolehannya berbeda. Dengan menggunakan teori tes modern informasi mengenai karakteristik tes lebih banyak diperoleh. Oleh sebab itu, dalam penelitian ini karakteristik tes dipaparkan berdasarkan teori tes modern. Setelah dilakukan penelitian dan dilakukan analisis berdasarkan teori tes modern. Setelah dilakukan penelitian dan dilakukan analisis berdasarkan teori respon butir diperoleh karakteristik dari ketiga perangkat tes seperti yang disajikan pada hasil penelitian. Pembahasan untuk masing-masing perangkat tes dapat diuraikan sebagai berikut :

Perangkat tes Matematika kelas VII

Secara klasik karakteristik butir Matematika kelas VII diperoleh data seperti pada Tabel 16. Berdasarkan tabel tersebut dapat disimpulkan bahwa perangkat tes Matematika kelas VII ada 7 butir yang gugur dimana korelasi biserialnya kurang dari 0,25, butir-butir tersebut adalah butir 1,4, 11,13, 16, 19 dan 25. Pada butir 1,4, dan 19 soalnya berbentuk perhitungan yang tergolong pada tingkat kesukaran mudah sehingga hampir semua siswa dapat menjawabnya, hal ini menyebabkan daya bedanya menjadi rendah. Butir 11,16, dan 25 bersifat hitungan yang mudah pula pemahaman terhadap materi persamaan sederhana dan harga, siswa dapat mengerjakan soal karena tidak terlalu sulit. Sedangkan butir 13 daya bedanya tinggi, hal ini kemungkinan disebabkan siswa sulit untuk memahami soal mengenai persamaan. Pada perangkat tes kelas VII yang digunakan dalam estimasi parameter butir adalah sebanyak 23 butir.

Berdasarkan hasil analisis dengan menggunakan teori respon butir dapat disimpulkan bahwa keseluruhan butir cocok

dengan model, karena memenuhi kriteria yang telah ditetapkan, sehingga keseluruhan butir dinyatakan lolos. Perangkat tes Matematika kelas VII memiliki daya pembeda secara keseluruhan adalah 0,673. Menurut Baker (2001), daya pembeda antara 0,65 sampai 1,34 dikatakan moderate. Oleh karena itu, daya pembeda sebesar 0,673 tersebut dikategorikan moderate (sedang), dengan daya pembeda seperti ini, butir-butir di dalam tes Matematika kelas VII diharapkan masih mampu membedakan peserta tes/individu yang memiliki kemampuan tinggi dan peserta tes/individu yang memiliki kemampuan rendah. Tingkat kesukaran butir tes Matematika kelas VII bergerak dari -0,107 sampai 1,112, dengan tingkat kesukaran rata-rata sebesar 0,608. Berdasarkan uraian tersebut dapat disimpulkan bahwa butir-butir soal dalam tes Matematika kelas VII tersebut memiliki karakteristik yang cukup memadai dari sudut statistik tes klasik dan tes modern, dalam arti cukup ideal dimasukkan ke dalam bank soal (Matematika).

Perangkat tes Matematika kelas VIII

Secara klasik karakteristik butir tes Matematika kelas VIII diperoleh data seperti pada Tabel 7. Berdasarkan tabel ini dapat disimpulkan bahwa perangkat tes Matematika kelas VIII ada 6 butir yang gugur dimana korelasi biserialnya kurang dari 0,25, butir-butir tersebut adalah butir 1, 2, 4, dan 18 mengenai soal persamaan dan fungsi kuadrat. Hal ini dalam pengerjaannya diperlukan tingkat pemahaman yang lebih dalam mengenai materi tersebut. Untuk menjawab soal nomor 20 siswa harus memahami konsep himpunan, butir 28 mengenai soal trigonometri. Oleh karena itu siswa harus memahami konsep tersebut siswa mengalami kesulitan. Pada perangkat tes kelas VIII yang digunakan dalam proses estimasi parameter adalah sebanyak 24 butir.

Hasil analisis yang berdasarkan teori tes modern dengan menggunakan program Bilog 3.00, menunjukkan bahwa keseluruhan butir memenuhi syarat dalam estimasi parameter butir serta ada 3 butir yang tidak cocok dengan model. Oleh karena itu ketiga butir tersebut tidak memenuhi kriteria yang telah ditetapkan, sehingga tidak dimasukkan ke dalam bank soal. Perangkat tes Matematika kelas VIII memiliki daya pembeda yang terentang dari 0,674 sampai 7,456 serta rata-rata daya pembeda secara keseluruhan adalah

1,524. Menurut Baker (2001), daya pembeda antara 0,65 sampai 1,34 dikatakan moderate. Oleh karena itu, daya pembeda sebesar 1,524 tersebut dikategorikan tinggi. Dengan daya pembeda seperti ini, butir-butir di dalam tes Matematika kelas VIII diharapkan mampu membedakan peserta tes/individu yang memiliki kemampuan tinggi dan peserta tes/individu yang memiliki kemampuan rendah. Tingkat kesukaran butir tes Matematika kelas VIII bergerak dari $-0,036$ sampai $2,215$ dengan tingkat kesukaran rata-rata sebesar $0,922$. Meskipun memiliki tanda positif (mengarah ke kanan) namun secara umum tingkat kesukarannya cenderung netral atau mendekati titik tengah (sedang).

Berdasarkan uraian tersebut di atas dapat disimpulkan bahwa butir-butir soal dalam tes Matematika kelas VIII tersebut memiliki karakteristik yang cukup memadai dari sudut statistik tes klasik dan tes modern, dalam arti cukup ideal dimasukkan ke dalam bank soal yang ditujukan untuk mengukur kemampuan tertentu (kemampuan Matematika).

Perangkat tes Matematika kelas IX

Secara klasik karakteristik butir tes Matematika kelas IX diperoleh data seperti pada Tabel 9. Berdasarkan tabel tersebut dapat disimpulkan bahwa perangkat tes Matematika kelas IX ada 3 butir yang gugur dimana korelasi biserialnya kurang dari $0,25$, butir-butir tersebut adalah butir 6, 15 dan 23. Butir 6, soalnya terlalu panjang sehingga siswa kesulitan untuk memahaminya. Pada butir 6, soalnya mengenai kesebangunan, kemungkinan siswa lupa terhadap pernyataan tentang teori tersebut. Sedangkan pada butir 15 dan 23, soalnya mengenai lingkaran dan bangun ruang, dimana soal menggunakan cerita panjang dengan menghubungkan dengan rumus yang akan digunakan untuk perhitungan, sehingga siswa lebih sulit untuk memahaminya. Terdapatnya 3 item pertanyaan yang gugur, sehingga pada perangkat tes kelas IX yang digunakan dalam proses estimasi parameter adalah sebanyak 27 butir.

Berdasarkan hasil analisis menurut teori tes modern dengan menggunakan program BILOG versi 3.00 menunjukkan ada 1 butir yang tidak memenuhi kriteria estimasi parameter dan 4 butir yang tidak cocok dengan model. Kelima butir tersebut tidak memenuhi kriteria yang telah ditetapkan, sehingga tidak

tidak dimasukkan ke dalam bank soal. Perangkat tes Matematika kelas IX memiliki daya pembeda secara keseluruhan adalah $0,604$.

Menurut Baker (2001), daya pembeda kurang dari $0,65$ dikatakan *low* (rendah). Sehingga daya pembeda sebesar $0,604$ tersebut dikategorikan *low* (rendah). Dengan daya pembeda seperti ini, butir-butir di dalam tes Matematika kelas IX diharapkan masih mampu membedakan peserta tes/individu yang memiliki kemampuan tinggi dan peserta tes/individu yang memiliki kemampuan rendah. Tingkat kesukaran butir tes Matematika kelas IX bergerak dari $-1,146$ sampai $2,215$, dengan tingkat kesukaran rata-rata sebesar $0,380$. Meskipun memiliki tanda positif (mengarah ke kanan) namun secara umum tingkat kesukarannya cenderung netral atau mendekati titik tengah.

Berdasarkan uraian tersebut dapat disimpulkan bahwa butir-butir soal dalam tes Matematika kelas IX tersebut memiliki karakteristik yang cukup memadai dari sudut statistik tes klasik dan tes modern, dalam arti cukup ideal dimasukkan ke dalam bank soal yang ditujukan untuk mengukur kemampuan tertentu (kemampuan Matematika).

Simpulan

Berdasarkan temuan hasil penelitian dan pembahasan sebelumnya dapat disimpulkan:

1. Menggunakan teori tes klasik, karakteristik butir soal matematika kelas VII rata-rata daya beda soal cukup baik dengan nilai rata-rata r_{bis} minimal $0,390$. Rerata daya pembeda soal (*biserial*) adalah $0,563$, menunjukkan bahwa tes memiliki daya pembeda baik. Karakteristik butir soal matematika kelas VIII memiliki tingkat kesukaran rerata test adalah $0,413$, termasuk tes dengan kategori tingkat kesukaran sedang. Rerata daya pembeda soal (*biserial*) adalah $0,565$, menunjukkan bahwa tes memiliki daya pembeda baik. Karakteristik butir soal matematika kelas IX tingkat kesukaran rerata test adalah $0,419$, termasuk tes dengan kategori tingkat kesukaran sedang. Rerata daya pembeda soal (*biserial*) adalah $0,562$, menunjukkan bahwa tes memiliki daya pembeda baik. Hasil ini menunjukkan bahwa soal-soal Matematika yang diujikan tahunpelajaran 2012/2013 sudah cukup baik

- tingkat kesukarannya. Berdasarkan tes klasik perangkat tes kelas VII sejumlah 29 butir, perangkat tes kelas VIII sejumlah 27 butir, dan perangkat tes kelas IX terdiri 30 butir.
- Menggunakan teori respon butir dengan model atau parameter (*Model Marginal Maximum Likelihood*) karakteristik butir soal kelas VII memiliki daya beda sebesar 0,673 dan tingkat kesukaran antara 1,107 – 1,112 dengan semua butir memiliki harga $X_{hitung} < X_{tabel}$ hal ini dikatakan cocok dengan model 2 parameter. Karakteristik butir soal kelas VIII terdapat 3 butir soal yang memiliki daya beda lebih dari 2 (syarat $0 < a < 2$; 2) dan tingkat kesukaran antara -0,036-1,564 dengan terdapat 1 butir soal (butir 8) yang memiliki harga $X_{hitung} > X_{tabel}$. Dalam hal ini untuk perangkat tes kelas VIII terdapat 4 butir yang tidak diikutkan dalam proses penyetaraan dan juga tidak dimasukkan ke dalam Bank Soal. Karakteristik butir soal kelas IX memiliki daya beda sebesar 0,604 dan tingkat kesukaran antara -1,146 – 2,215 (terdapat 1 butir yang tidak memenuhi) dengan 1 butir memiliki harga $X_{hitung} > X_{tabel}$. Dalam hal ini untuk perangkat tes kelas IX terdapat 2 butir yang tidak diikutkan dalam proses penyetaraan dan juga tidak dimasukkan ke dalam Bank Soal.
 - Perbandingan hasil analisis butir soal model teori tes klasik dengan model teori tes modern tidak terlalu signifikan, hal ini bisa disebabkan jumlah peserta tes kurang dari 100 dan selisihnya hanya 4-6 siswa. Hasil kajian menunjukkan, pada jumlah peserta ujian yang banyaknya 60 orang, jumlah atau persentase butir soal yang valid dengan menggunakan pendekatan modern sama banyak dengan pendekatan klasik. Sedangkan pada jumlah peserta ujian yang banyaknya kurang dari 60 orang, jumlah atau persentase butir soal dengan menggunakan pendekatan tes klasik lebih banyak.
 - Estimasi kemampuan peserta ujian mata pelajaran Matematika kelas VII, VIII, dan IX dengan menggunakan model *Marginal Maximum Likelihood* cukup beragam. Pada tahun pelajaran 2012/2013 kisaran estimasi kemampuan peserta ujian sama yakni berkisar antara 0,2650 sampai dengan 0,5834. Hal ini menunjukkan bahwa kemampuan peserta ujian Matematika kelas VII (P), VIII (Q), dan IX (R) pada tahun pelajaran 2012/2013 masih rendah. Hal ini dapat disebabkan karena peserta ujian tidak mampu mengerjakan secara maksimal yang bisa disebabkan karena kurangnya persiapan dalam belajar.
 - Persamaan atau formula konversi skor dari tiga perangkat tes mata pelajaran matematika Sekolah Menengah Pertama Negeri 1 Tidore Kepulauan, yaitu antara perangkat tes P dan perangkat tes Q adalah, $P^* = 1,084P - 0,221$ dan $Q^* = 0,922Q + 2,965$. Persamaan antara perangkat tes dan R adalah, $Q^* = 0,922Q + 2,965$ dan $R^* = 0,998R + 2,775$. Persamaan antara perangkat tes P dan R' adalah, $P^* = 1,084P - 0,221$ dan $R^* = 0,998R + 2,775$. Dari ketiga perangkat tes tersebut antara P dan Q lebih sulit perangkat tes Q, antara perangkat tes Q dan R lebih sulit perangkat tes Q, antara perangkat tes P dan R lebih sulit perangkat tes P. Penyetaraan perangkat tes kelas VII ke kelas VIII mengalami penurunan, menjadi lebih mudah bila dikerjakan oleh siswa kelas VIII. Begitu pula penyetaraan perangkat tes kelas VIII ke kelas IX mengalami penurunan, menjadi lebih mudah bila dikerjakan oleh siswa kelas IX. Namun Penyetaraan perangkat tes kelas VII ke kelas IX mengalami peningkatan, menjadi lebih sulit bila dikerjakan oleh siswa kelas IX. Hal ini bisa disebabkan karena kelas IX kurang memahami kembali pelajaran kelas VII yang sudah terlewati.

Saran

- Penyetaraan perangkat tes penting untuk dilakukan dalam penilaian hasil belajar, pada ujian akhir semester perlu mengembangkan perangkat tes yang setara untuk setiap mata pelajaran. Oleh karena itu, baik guru maupun pihak-pihak yang terlibat dalam pengembangan tes, dianjurkan untuk mengupayakan beberapa perangkat tes yang telah disetarakan. Perangkat tes dengan persamaan konversi dapat digunakan untuk menilai hasil belajar siswa atau peserta tes. Selain itu perangkat tes juga dapat digunakan untuk memantau kemajuan prestasi siswa pada suatu daerah dari waktu ke waktu.
- Ujian tertulis Matematika untuk kelas yang

memiliki peserta cukup banyak, sebaiknya menggunakan pendekatan tes yang mengarah pada tes modern. Pendekatan tes klasik masih cukup memadai untuk ujian Matematika yang pesertanya masih sedikit seperti halnya pada kelas IX.

3. Penelitian ini memiliki kelemahan tidak mengkaji hasil analisis butir soal dengan kisi-kisi dan lembar indikator soal setiap kelas. Penelitian selanjutnya disarankan kajian yang komprehensif dengan menggabungkan hasil analisis butir soal dan dokumen soal yang terkait seperti soal ujian, kisi-kisi, dan lembar indikator.

Daftar Pustaka

- Baker, F. B. (2001), *The of Item Response Theory*, pp. 456, New York: Cleringhouse on Assesment and Evaluation.
- Crocker, L. & Algina, J. (1986), *Introduction To Classical and Modern Test Theory*, pp. 456, New York: University of Florida.
- Depdiknas (2005), *Peraturan Pemerintah No 19, Standar Nasional Pendidikan*, Jakarta: Depdiknas.
- Hambleton, R.K. Swaminathan, H. & Rogers, H.J (1985), *Item Response Theory*, Boston: Kluwer Nijjh off Publisher.
- Kemendiknas, (2011). *Modul Matematika SMP Program Bermutu*, Yogyakarta: Kemendiknas.
- Kolen, M.I., dan Brennan, R.I. (1995), *Test Equating: Method and Pratices*, pp. 14, Iowa: University of Iowa.
- Linn, (1989), *Educational Measurement* (3rd ed), New York: Macmillan Publising Company.
- Suryabrata, S. (1998). *Pengembangan Alat Ukur Psikologis*. Jakarta: Ditjen Dikti, Depdikbud.